

**ONLINE TOOL TO DETECT PLAGIARIZED STUDENTS'
SOLUTIONS FOR CLASSES' ASSIGNMENTS**

Submitted by

Hisham S.M. Khdair (801174)

©2010



KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

HISHAM S.M KHDAIR
(801174)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**


telah mengemukakan kertas projek yang bertajuk
(has presented his/ her project paper of the following title)

ONLINE TOOL TO DETECT PLAGIARIZED STUDENTS' SOLUTIONS FOR
CLASSES' ASSIGNMENTS

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan
dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory
knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **DR.AZIZAH BT HAJI AHMAD**

Tandatangan
(Signature) : 

Tarikh
(Date) : 19 May 2010

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of the Graduate School.

It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis. Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Dean of Postgraduate

College of Arts and Sciences (UUM-CAS)

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman.

ABSTRACT

Most of the existing plagiarism detection systems compare documents to archived sources in enterprise databases. The aim of this study is to develop an online tool to detect plagiarism and cheating for students' assignments focusing on those who copy from each others'. This tool provide an easy and automated way to enable lecturers to evaluate students' solution for a given assignments. Online Detecting of Assignments' Plagiarism (ODAP) integrates the Longest Common Substring algorithm (LCS) for comparing strings of documents. ODAP has been tested and evaluated by academics and students of Msc(Information Technology) of UUM, and it has wide acceptance and satisfaction of its functionalities and outcomes.

Acknowledgement

**IN THE NAME OF ALLAH, THE MOST GRACIOUS AND MOST
MERCIFUL**

First and foremost, all praise to Allah for providing me with the strength, perseverance, and wisdom to have this work done on time.

I would like to thank and extend my deep gratitude to my supervisor **DR. AZIZAH BT HAI AHMAD**. I was extremely fortunate to have her as my primary source of knowledge, guidance, and support throughout the process of my project, and I would certainly be even more fortunate to benefit for her wide experience in my future academic undertakings. In short, this work could not have been accomplished without her enlightenment. Thank you very much.

Special thank to those who supported me with their prayers particularly my mother, aunt, and all of my brothers and sisters. I would like also to thank my friends in my homeland Palestine and also in Malaysia all by his name. I won't forget to dedicate this work to the soul of my father may ALLAH bless his soul and make his grave a garden and grant him the highest levels of paradise... Ameen.

Hisham S.M. Khdair

May 5th 2010

TABLES OF CONTENTS

| | |
|---|-----------|
| CHAPTER 1: INTRODUCTION..... | 1 |
| 1.1 BACKGROUND..... | 1 |
| 1.2 PROBLEM STATEMENT..... | 2 |
| 1.3 RESEARCH AIM..... | 3 |
| 1.4 RESEARCH OBJECTIVES..... | 3 |
| 1.5 SCOPE AND LIMITATION..... | 4 |
| 1.6 SIGNIFICANCE OF THE STUDY..... | 4 |
| 1.7 ORGANIZATION OF THE REPORT..... | 5 |
| CHAPTER 2: LITERATURE REVIEW..... | 6 |
| 2.1 PLAGIARISM..... | 6 |
| <i>2.1.1 CONVENTIONAL WAYS OF PLAGIARISM.....</i> | <i>7</i> |
| 2.2 CONSEQUENCES OF PLAGIARISM..... | 11 |
| <i>2.2.1 UNDERSTANDING WHY PLAGIARISM IS DISLIKED AND ODIOUS.....</i> | <i>12</i> |
| 2.3 HANDLING PLAGIARISM..... | 12 |
| 2.4 PLAGIARISM DETECTION..... | 14 |
| <i>2.4.1 PLAGIARISM DETECTION SYSTEMS.....</i> | <i>15</i> |
| 2.4.1.1 SEARCH ENGINES..... | 15 |
| 2.4.1.2 COMMERCIAL SYSTEMS (E.G. TURNITIN)..... | 16 |
| 2.4.1.3 PROGRAMMING CODE SIMILARITY DETECTION SYSTEMS..... | 17 |
| <i>2.4.2 PLAGIARISM DETECTION ALGORITHMS.....</i> | <i>18</i> |
| 2.4.2.1 LONGEST COMMON SUBSTRING..... | 18 |
| 2.4.2.2 LONGEST COMMON SUBSEQUENCE..... | 20 |

| | |
|---|-----------|
| 2.5 SUMMARY..... | 21 |
| CHAPTER 3: RESEARCH METHODOLOGY | 22 |
| 3.1 AWARENESS OF PROBLEM..... | 23 |
| 3.2 SUGGESTION..... | 23 |
| 3.3 DEVELOPMENT..... | 23 |
| 3.4 EVALUATION..... | 25 |
| 3.5 CONCLUSION..... | 25 |
| CHAPTER 4: SYSTEM ANALYSIS AND DESIGN..... | 27 |
| 4.1 SYSTEM REQUIREMENTS..... | 27 |
| 4.1.1 FUNCTIONAL REQUIREMENTS..... | 27 |
| 4.1.2 NON-FUNCTIONAL REQUIREMENT..... | 29 |
| 4.2 SYSTEM DESIGN..... | 29 |
| 4.2.1 USE CASE DIAGRAM..... | 30 |
| 4.2.2 SYSTEM FUNCTIONALITIES (SEQUENCE DIAGRAM)..... | 31 |
| 4.2.2.1 VIEW ASSIGNMENT SEQUENCE DIAGRAM..... | 31 |
| 4.2.2.2 UPLOAD ASSIGNMENT SOLUTION SEQUENCE DIAGRAM.... | 32 |
| 4.2.2.3 MANAGE ASSIGNMENTS SEQUENCE DIAGRAM..... | 34 |
| 4.2.2.4 VIEW SOLUTIONS SEQUENCE DIAGRAM..... | 35 |
| 4.2.3 CLASS DIAGRAM..... | 36 |
| 4.3 PROTOTYPE DESIGN..... | 38 |
| 4.3.1 STUDENT'S INTERFACES..... | 38 |
| 4.3.1.1 STUDENT LOGIN INTERFACE..... | 38 |
| 4.3.1.2 VIEW COURSE INTERFACE..... | 39 |
| 4.3.1.3 VIEW ASSIGNMENTS INTERFACE..... | 40 |

| | |
|--|-----------|
| 4.3.1.4 VIEW AN ASSIGNMENT AND SUBMIT SOLUTION INTERFACE..... | 41 |
| 4.3.1.5 VIEW/DOWNLOAD AND THE SUBMITTED SOLUTION..... | 42 |
| 4.3.2 LECTURER'S INTERFACES..... | 43 |
| 4.3.2.1 LECTURER LOGIN INTERFACE..... | 43 |
| 4.3.2.2 VIEW COURSES INTERFACE..... | 44 |
| 4.3.2.3 ASSIGNMENTS INTERFACE..... | 45 |
| 4.3.2.4 VIEW STUDENTS' SUBMISSION..... | 46 |
| 4.3.2.5 SIMILARITY REPORT FORM..... | 47 |
| 4.4 ALGORITHM IMPLEMENTATION..... | 49 |
| 4.5 SUMMARY..... | 51 |
| CHAPTER 5: FINDING | 52 |
| 5.1 USABILITY EVALUATION..... | 52 |
| 5.2 INSTRUMENT FOR USER EVALUATION..... | 52 |
| 5.3 GENERAL INFORMATION..... | 53 |
| 5.4 RELIABILITY ANALYSIS TEST..... | 54 |
| 5.5 DESCRIPTIVE STATISTICS FOR ALL ITEMS..... | 55 |
| 5.6 SUMMARY..... | 59 |
| CHAPTER 6: CONCLUSION | 60 |
| 6.1 CONTRIBUTION OF THE STUDY..... | 60 |
| 6.2 LIMITATION AND FUTURE WORK..... | 61 |
| REFERENCES..... | 62 |
| APPENDIX A: QUESTIONNAIRE..... | 66 |

LIST OF TABLES

TABLE 2-1: METHODS OF PLAGIARISM STUDENTS USE IN THEIR CLASS'S ASSIGNMENTS
AND WORK8

TABLE 2-2: PLAGIARISM AND THE INTERNET 10

TABLE 2-3: ATTRIBUTES OF DETECTION AND PREVENTION METHODS..... 13

TABLE 3-1: OBJECTIVES AND FOLLOWED METHODS TO VERIFY ITS OUTCOMES25

TABLE 4-1: FUNCTIONAL REQUIREMENTS OF ODAP28

TABLE 4-2: LIST OF NON-FUNCTIONAL REQUIREMENTS29

TABLE 5-1: PROFILE OF QUESTIONNAIRE 'S RESPONDENTS53

TABLE 5-2: RESULTS OF RELIABILITY TEST.....54

TABLE 5-3: STATISTICS OF RESPONSES FOR COMPUTER SYSTEM USABILITY
QUESTIONNAIRE55

LIST OF FIGURES

| | |
|---|----|
| FIGURE 2-1: ETBLAST: A SIMILARITY-BASED SEARCH ENGINE..... | 15 |
| FIGURE 2-2: SCREENSHOT OF TURNITIN GENERATED PLAGIARISM REPORT | 16 |
| FIGURE 2-3: MULIPLE STRINGS REPRESENT AS AN INPUT TO LCS ALGORITHM | 19 |
| FIGURE 2-4: RESULTED LONGEST SUBSTRINGS RETURNED OUT FROM LCS ALGORITHM | 19 |
| FIGURE 2-5: FILE OUTPUT USING DIFF IMPLEMENTING LONGEST COMMON SUBSEQUENCE | 20 |
| FIGURE 3-1: GENERAL METHODOLOGY OF INFORMATION SYSTEM'S RESEARCHES..... | 22 |
| FIGURE 3-2: SOFTWARE DEVELOPMENT LIFE CYCLE (PROTOTYPING MODEL) | 24 |
| FIGURE 4-1: USE CASE DIAGRAM..... | 30 |
| FIGURE 4-2: VIEW ASSIGNMENT SEQUENCE DIAGRAM..... | 32 |
| FIGURE 4-3: UPLOAD ASSIGNMENT SOLUTION SEQUENCE DIAGRAM..... | 33 |
| FIGURE 4-4: MANAGE ASSIGNMENTS SEQUENCE DIAGRAM | 34 |
| FIGURE 4-5: VIEW SOLUTIONS SEQUENCE DIAGRAM..... | 35 |
| FIGURE 4-6: SYSTEM CLASS DIAGRAM..... | 37 |
| FIGURE 4-7: STUDENT LOGIN INTERFACE | 38 |
| FIGURE 4-8: VIEW COURSE INTERFACE..... | 39 |
| FIGURE 4-9: VIEW ASSIGNMENTS INTERFACE | 40 |
| FIGURE 4-10: VIEW AN ASSIGNMENT AND SUBMIT SOLUTION INTERFACE..... | 41 |
| FIGURE 4-11: VIEW/DOWNLOAD AND THE SUBMITTED SOLUTION | 42 |
| FIGURE 4-12: VIEWING THE UPLOADED ASSIGNMENT IN HTML FORM..... | 43 |
| FIGURE 4-13: LECTURER LOGIN INTERFACE | 44 |
| FIGURE 4-14: VIEW COURSES INTERFACE | 44 |
| FIGURE 4-15: PAGE OF AVAILABLE ASSIGNMENTS..... | 45 |
| FIGURE 4-16: VIEW STUDENT'S UPLOADED ASSIGNMENT..... | 46 |

FIGURE 4-17: CHECKING SIMILARITY FOR ALL STUDENT REPORT FORM47

FIGURE 4-18: DETAILED TEXT SIMILARITY REPORT48

FIGURE 5-1: STUDENTS' RESPONSES FOR EASE OF USE ITEMS.....57

FIGURE 5-2: LECTURERS' RESPONSES FOR EASE OF USE ITEMS.....58

FIGURE 5-3: STUDENTS' RESPONSES FOR USEFULNESS ITEMS.....58

FIGURE 5-4: LECTURERS' RESPONSES FOR USEFULNESS ITEMS.....59

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Assignments are so essential for educators in evaluating students' work throughout class life. The submission of assignments by computers opens the chance to students, and eases the way to copy each others solution (Harris, 2001). Many students duplicate their answers in their homework by copying and pasting parts or even entire solution.

Plagiarism is the “wrongful act of taking the product of another person’s mind, and presenting it as one’s own” (Gibaldi, 1999). It is considered as a growing problem for the scientific society in general and universities in particular (Clough, 2000b; Culwin & Lancaster, 2000). This trend is appeared to be prevalent and in so many countries. Including US (Park, 2003), Finland (Seppanen, 2002), and South Africa (Weeks, 2001).

Intentional or substantial plagiarism however, may be grounds for failing an assignment or class, receiving disciplinary action, and being suspended or dismissed from the university (Gibaldi, 1999).

Beyond that, a survey of 149 of undergraduates in the US, concluded that 87% of students thought that, those who used to cheat in the academic life will be more likely to do such behavior again in work (Morgan & Foster, 1992). Universities ignoring or even tolerating the problem imperil their reputation for honesty and fairness. Therefore, detecting the plagiarism of students' assignments among each others is a major concern for all lecturers. Lecturer requires to find those students who cheat and

The contents of
the thesis is for
internal user
only

REFERENCES

- Alla Anohina, A. M., & Grundspenkis, J. (2009). Development of the Plagiarism Detection Tool for Processing Template-Based Documents.
- Ashworth, P., Bannister, P., & Thorne, P. (1997). Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment.
- Barclay, K., & Savage, J. (2004). Object-Oriented Design with UML and Java.
- Bergroth, L., Hakonen, H., & Raita, T. (2000). A survey of longest common subsequence algorithms.
- Carroll, J., & Appleton. (2001). Plagiarism - A Good Practice Guide. Technical report.
- Carroll, J., & Seymour, D. (2006). The effect of a penalty tariff on consistent decision-making in cases of student plagiarism.
- Clough, P. (2000b). Analysing style - Readability. Technical report, University of Sheffield.
- Coakes, S. J. (2005). SPSS version 12 for Windows Analysis Without Anguish. Sydney: John Wiley & Sons Australia.
- Culwin, F., & Lancaster, T. (2000). A Review of Electronic Services for Plagiarism Detection in Student Submissions. Paper presented at the Proceedings of the 1st LTSN-ICS.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *International Journal of Human-Computer Interaction*, 7(1), 57-78.
- Day, F. R., & William, H. E. (1993). The computation of consensus patterns in DNA sequences. *Mathematical and Computer Modelling*, 17(10), 49-52.
- Dey, S. K., & Sobhan, M. A. (2006). Impact of Unethical Practices of Plagiarism on Learning, Teaching and Research in Higher Education: Some Combating Strategies.
- East, J. (2009). Judging plagiarism: a problem of morality and convention. *Higher Education*, 59(1), 69-83.
- Eriksson, H. E., & Penker, M. (2000). Business modeling with UML: Wiley.

- Evans, R. (2006). Evaluating an electronic plagiarism detection service: The importance of trust and the difficulty of proving students don't cheat. *Active Learning in Higher Education*, 7(1), 87.
- Gibaldi, J. (1999). *MLA Handbook for Writers of Research Papers* (5th ed.).
- Gingerich, J. (2010). AV EX. REL. VANDERHYE V. IPARADIGMS, llc: electronic databases and the compartmentalization of fair use.
- Gruner, S., & Naven, S. (2005). Tool Support for Plagiarism Detection in Text Documents. *ACM Symposium on Applied Computing*.
- Harris, R. (2001). *The Plagiarism Handbook: Strategies for Preventing, Detecting, and Dealing with Plagiarism*: Pyczak Publishing.
- Johan, 2004 "information system analysis and design" retrieved: October 2005, available from: <http://www.cs.toronto.edu/~jm/3405/slides2/sequence D.pdf>.
- Jordan, A. E. (2001). College student cheating: The role of motivation, perceived norms, attitudes, and knowledge of institutional policy. *Ethics & Behavior*, 11(3), 233-247.
- Joy, M., & Luck, M. (1999). Plagiarism in Programming Assignments. *IEEE TRANSACTIONS ON EDUCATION*.
- Kiehl, E. M. (2006). Using an ethical decision-making model to determine consequences for student plagiarism. *Journal of Nursing Education*, 45(6), 199.
- Lancaster, T., & Culwin, F. (2005). CLASSIFICATIONS OF PLAGIARISM DETECTION ENGINES.
- Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1), 57-78, available from: <http://oldwww.acm.org/perlman/question.cgi#q4>.
- Liddell, J. (2003). A Comprehensive Definition of Plagiarism. *Community & Junior College Libraries*, 11(3), 43-52.
- Lindey, A. (1952). *Plagiarism and Originality*.
- Lukashenko, R., Graudina, V., & Grundspenkis, J. (2007). Computer-based plagiarism detection methods and tools: an overview.
- Macdonald, R., & Carroll, J. (2006). Plagiarism—a complex issue requiring a holistic institutional approach. *Assessment & Evaluation in Higher Education*, 31(2), 233-245.
- Magel, K. (1981). Regular expressions in a program complexity metric. *ACM*.

- Martin, B. (1994). Plagiarism: a misplaced emphasis. *Journal of Information Ethics & Behavior*.
- McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11(3), 219-232.
- McGowan, U. (2005). Plagiarism detection and prevention: Are we putting the cart before the horse?
- Morgan, C. J., & Foster, W. T. (1992). Student Cheating: An ethical dilemma. Paper presented at the ASEE, Frontiers in Education Conference.
- Nakatsu, N., Kambayashi, Y., & Yajima, S. (1982). A longest common subsequence algorithm suitable for similar text strings. *Acta Informatica*, 18(2), 171-179.
- Naumann, J. D., & Jenkins, A. M. (1982). Prototyping: the new paradigm for systems development. *MIS Quarterly*, 6(3), 29-44.
- Pa, J. (2008). PLAGIARISM DETECTION TECHNIQUES.
- Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS* (3rd ed.). Wellington, New Zealand: Allen and Unwin.
- Park, C. (2003). In other (people's) words: plagiarism by university students – literature and lessons. *Assessment and Evaluation in Higher Education*.
- Phillips, R. (2002). Plagiarism and Theological Education. *Journal of Religious & Theological Information*, 5(2), 3-12.
- Rahmann, S. (2003). Fast large scale oligonucleotide selection using the longest common factor approach. *INTERNATIONAL JOURNAL OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY*, 1(2), 343-362.
- Rambally, G. K., & Sage, M. L. (1990). An inductive inference approach to plagiarism detection in computer programs.
- Romans Lukashenko, Graudina, V., & Grundspenkis, J. (2007). Computer Based Plagiarism Detection Methods and Tools an Overview. Paper presented at the Proceedings of the 2007 international conference on Computer systems and technologies.
- Roy, C. K., & Cordy, J. R. (2007). A survey on software clone detection research. *Queen's School of Computing TR*, 541, 115.
- Scanlon, P. M. (2003). Student Online Plagiarism: How Do We Respond? *College Teaching*, 51(4), 161-165.
- Scanlon, P. M., & Neumann, D. R. (2002). Internet plagiarism among college students. *Journal of College Student Development*, 43(3), 374-385.

- Seppanen, R. (2002). Finns target master plagiarists.
- Smith, A. G., & Macintosh, B. (1999). Penalties for Plagiarism.
- Vaishnavi, V., & Kuechler, W. (2004/2005). Design Research in Information Systems.
- Weeks, S. (2001). Plagiarism: think before pointing finger of blame.
- White, E. M. (1993). Too many campuses want to sweep student plagiarism under the rug. *Chronicle of Higher Education*, 24, A 44.
- Wieggers, Karl E. (2003). *Software Requirements 2: Practical techniques for gathering and managing requirements throughout the product development cycle* (2nd ed.). Redmond: Microsoft Press. ISBN 978-0735618794.
- Worthen, K. J. (2004). Discipline: An Academic Dean's Perspective on Dealing with Plagiarism. *BYU Educ. & LJ*, 441.
- Zikmund, W.G. (2000) *Business research Methods* (3 th ed.), fort worth: Harcourt College Publishers.
- Zobel, J., & Moffat, A. (2006). Inverted files for text search engines. *ACM Computing Surveys*